Article

Chinese Science Bulletin

Progress of Projects Supported by NSFC
Hydraulic Engineering

September 2012 Vol.57 No.26: 3397–3403
doi: 10.1007/s11434-012-5305-x

Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model

SONG XiaoMeng[1], ZHAN CheSheng[2*] & XIA Jun[2,3]

[1] *Nanjing Hydraulic Research Institute, Nanjing 210029, China;*
[2] *Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;*
[3] *State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China*

Received October 26, 2011; accepted May 7, 2012

Parameter optimization of a hydrological model is an indispensable process within model development and application. The lack of knowledge regarding the efficient optimization of model parameters often results in a bottle-neck within the modeling process, resulting in the effective calibration and validation of distributed hydrological models being more difficult to achieve. The classical approaches to global parameter optimization are usually characterized by being time consuming, and having a high computation cost. For this reason, an integrated approach coupling a meta-modeling approach with the SCE-UA method was proposed, and applied within this study to optimize hydrological model parameter estimation. Meta-modeling was used to determine the optimization range for all parameters, following which the SCE-UA method was applied to achieve global parameter optimization. The multivariate regression adaptive splines method was used to construct the response surface as a surrogate model to a complex hydrological model. In this study, the daily distributed time-variant gain model (DTVGM) applied to the Huaihe River Basin, China, was chosen as a case study. The integrated objective function based on the water balance coefficient and the Nash-Sutcliffe coefficient was used to evaluate the model performance. The case study shows that the integrated method can efficiently complete the multi-parameter optimization process, and also demonstrates that the method is a powerful tool for efficient parameter optimization.

parameter optimization, statistical emulator approach, response surface method, SCE-UA, distributed hydrological model, Huaihe River basin

Citation: Song X M, Zhan C S, Xia J. Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model. Chin Sci Bull, 2012, 57: 3397–3403, doi: 10.1007/s11434-012-5305-x

Hydrological models are the principal tools used to investigate hydrological processes within watersheds [1]. Distributed hydrological models have become increasingly popular for hydrology research and water resources management. Usually, these hydrological models are conceptual models that simplify and interpret actual hydrological processes using a mathematical formula and physical equations. The reliability of model predictions depends on how rigorously the model structure is defined and how rigorously the model is parameterized [2]. However, the accurate estimation of model parameters is difficult due to the large uncertainties involved, as parameters usually cannot be directly measured in the field, or their exact values are not known [3,4]. Therefore, parameter calibration and optimization becomes necessary to improve model performance within most model applications [5]. During the model calibration and optimization process, selected parameters are allowed to vary within predefined bounds until an optimization objective is met. There is also the problem of equifinality. Models that use a large number of parameters can have multiple combinations of parameter values that give suitable predictions of observed data. However, when a large number of parame-

*Corresponding author (email: zhancs2006@gmail.com)

ters exist within a model, the calibration process can become complex, time-consuming and is characterized by high computation cost. In addition, high nonlinearity and multimodality of models due to a large number of parameters used, results in some optimization issues which require complex solutions [6]. In many instances, complex optimization problems include noise and/or discontinuities within the set of Pareto solution data, making traditional deterministic methods inefficient for finding global solutions. Global optimization methods based on meta-heuristics are robust alternatives for solving complex optimization problems [6]. Over the last several decades, there has been an increasing interest in meta-heuristics as well as improvements in its application to optimization problems. For example the Shuffled Complex Evolution method at the University of Arizona (SCE-UA) [7], simulated annealing [8], tabu search [9], genetic algorithms [10], differential evolution [11], particle swarm optimization [12], ant colony optimization [13] and scatter search [14].

However, the application of system analytical techniques for computationally demanding models, such as optimization, sensitivity analysis, and statistical inference, may be hampered by the high computational cost associated with multiple evaluations of a model [15]. The computational cost for optimization is enormous, and this poses a major problem to hydrologists, even with the great improvements in computational algorithms and computing hardware. For this reason, an efficient and effective optimization method should be developed to reduce the computation demand or cost. There are two strategies for dealing with this problem, namely, improving the efficiency of the model evaluation (e.g. simplifying the model if possible), and improving the efficiency of the computationally demanding techniques [15]. The computationally demanding model can be replaced by an efficient emulator of the model, i.e. the meta-modeling approach can be used to evaluate and optimize complex model parameters [15–20]. There exist a variety of meta-modeling techniques [19], including polynomial regression, Kriging modeling, multivariate adaptive regression splines (MARS), radial-basis functions (RBF), multi-layer perception networks (MLP), and support vector machines (SVM). Usually, the meta-models are a statistical approximation of deterministic models [21]. They can in principle be applied to dynamic models as well. However, they have two significant deficiencies [15], namely the lack of a priori cognition of the model structure, and numerical difficulties resulting from a large number of closely spaced input points within the dynamic models with a multi-dimensional temporal scale, especially with regard to hydrological models.

Given that a complex hydrological model is computationally intensive and that the meta-modeling approach may amplify the uncertainties or errors, the integrated method, coupling a meta-modeling approach with a classical optimization method was employed for parameter optimization, to meet the desired objective in this study. In this study, sensitivity and uncertainty analysis identified important or sensitive parameters as well as non-sensitive or unimportant parameters. Important parameters were optimized and unimportant parameters were fixed within the optimization process. The response surface method was used to estimate the value of parameters, following which the ranges of parameters were adjusted to obtain relatively reliable ranges and bounds. Finally, the SCE-UA method was used in the optimization process based on the adjusted parameter ranges.

1 Methodology

1.1 SCE-UA method

The SCE-UA method is a global searching algorithm proposed by Duan et al. [7,22] and it has been used in many hydrological models. The SCE-UA method combines the direction-searching of deterministic, non-numerical methods and the robustness of stochastic, non-numerical methods. It adopts competition evolution theory, concepts of controlled random search, the complex shuffling method, and downhill simplex procedures to obtain a global optimal estimation [7]. To make the implementation more convenient, Duan suggested some default values for the parameters of the SCE-UA method [7,22].

1.2 Statistical emulator approach

The statistical emulator approach, which has been widely used within uncertainty analysis and parameter optimization, is a collection of statistical and mathematical techniques useful for developing, improving, and optimizing processes. The response surface method (RSM) is a common statistical emulator method, and was introduced by Box and Wilson in 1951 [23], which explored the statistical relationships between several explanatory variables and one or more response variables. The main idea of RSM is to use a sequence of design experiments to obtain an optimal response. In this study, the PSUADE (Problem Solving environment for Uncertainty Analysis and Design Exploration) developed by Lawrence Livermore National Laboratory was used to construct the response surface between the model parameters and objective functions using the MARS method. The MARS is a non-parametric regression technique introduced by Friedman in 1991 [24], and can be regarded as an extension of linear models that automatically simulate non-linear relationships and interactions. Compared with the neural network and Gaussian process, the MARS technique has been particularly popular because it does not assume or impose any particular type or class of relationship (e.g. linear or logistic) between the predictor variables and the dependent variable of interest. Instead, it allows the regression function to be driven directly by the data [25]. It is also more flexible than the linear regression models, and easier to understand and interpret. More information about this technique and how to use it is included in the work of Friedman [24].

1.3 Integration of the RSM and the SCE-UA method

An integration method based on the RSM and SCE-UA method (RSM-SCE-UA) is proposed to estimate and optimize the parameters of hydrological models based on the PSUADE platform developed by the Lawrence Livermore National Laboratory. In this study, the RSM was used to calibrate the hydrological model and the SCE-UA method was applied to optimize the parameters. The implementation of the integration method involves the following steps as shown in Figure 1: (1) Determine what output is optimized, what input parameters should be adjusted, and what the ranges of the input parameters should be; (2) Choose an appropriate design of experiments for generating the parameter samples based on the PSUADE platform; (3) Run the DTVGM model to obtain the response objective functions for the samples; (4) Choose an approximation emulator to generate the RSM; (5) Examine the RSM to ensure that a global optimization value (minimum or maximum) lies on the response surface, and redefine the optimization ranges of parameters; (6) Run the SCE-UA method to optimize the parameters around the minimum or maximum on the response surface.

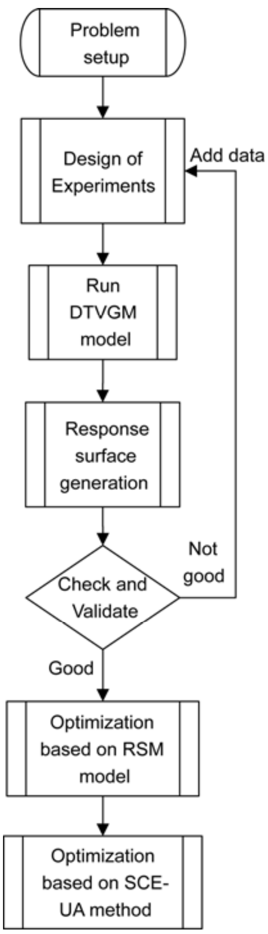

Figure 1 Flowchart showing the sequences within the RSM-SCE-UA method.

2 Study area and hydrological model

The Huaihe River (112°–121°E, 31°–36°N) is one of the seven longest rivers in China. It is located between the Yangtze River and the Yellow River. The river basin covers an area of 270000 km^2 and is inhabited by ~150 million people. The Huaihe River Basin has a fairly complicated water system, with a large number of tributaries, inter-provincial rivers, and artificial rivers used for water control. The Huaihe River Basin is divided into two major systems, i.e. the Huaihe River System and the Yi-Shu-Si River System. The average annual rainfall for the basin is about 900 mm, of which 70%–80% occurs as summer rainfall. The Huaihe River has an average annual flow of 853 $m^3 s^{-1}$, with maximum flood discharge >11000 $m^3 s^{-1}$ and nearly zero during the dry season. In this study, six years of data from 2003–2008, including the daily precipitation, evaporation, and discharge from the weather stations and hydrological stations was used to construct the DTVGM model for application to the Huaihe River basin, and to calibrate and optimize the model parameters. The description and theory of the DTVGM model are available in [26,27]. We selected the important or sensitive parameters based on the sensitivity analysis results to reduce the computation runs. The ranges of parameters are shown in the Table 1.

Regarding calibration and optimization, the performance of models can be evaluated in terms of the statistical measure of goodness-of-fit. In this paper, two objective functions (water balance coefficient WB and Nash-Sutcliffe efficiency coefficient NS) were selected as evaluation criterion to form an integrated objective function [28], as shown in eq. (1):

$$\mathrm{OBF} = w\left|1 - WB\right| + (1-w)\left|1 - NS\right| , \tag{1}$$

$$WB = \sum_{i=1}^{n} Q_{s,i} \Big/ \sum_{i=1}^{n} Q_{o,i} , \tag{2}$$

Table 1 Parameters of the DTVGM model and their ranges

Parameter	Description	Ranges
g_1	Time-variant gain factor, related to surface runoff generation	[0.01,1.0]
g_2	Time-variant gain factor, related to soil moisture content	[0.01,5.0]
Kr	Storage-outflow coefficient related to interflow runoff generation	[0.01,1.0]
Kaw	Coefficient for actual evapotranspiration	[0.01,1.0]
Wmi	Minimum soil moisture storage	[0.01, 0.40]
WM	Upper layer saturated soil moisture storage	[0.40,1.0]

$$NS = 1 - \sum_{i=1}^{n}\left(Q_{o,i} - Q_{s,i}\right)^2 \Big/ \sum_{i=1}^{n}\left(Q_{o,i} - \bar{Q}_o\right)^2, \quad (3)$$

where w is the weight of water balance (the default value is 0.5 [28]), Q_s and Q_o are the simulated value and observed value for the runoff, respectively, and $\bar{Q}$ is the mean value for corresponding runoff. Eq. (1) shows that the OBF can not only achieve a good agreement between the simulated and observed runoff volume, but also obtain a good simulation for hydrological processes.

3 Results and discussion

Experimental design and sampling design are crucial steps within the construction of a response surface model when we use the RSM. In this study, a space-filling design method (LP-τ method, also known as quasi-random sequence) [29] was used to generate the response surface. We used the PSUADE platform to generate the 4000 runs of parameters (2000 samples were used to construct the response surface, and the other 2000 samples were used to validate the model), and applied the DTVGM model to the Huaihe River Basin. The two objective functions WB and NS were calculated using the simulated discharge and observed discharge according to eqs. (2) and (3), following which the OBF was calculated using eq. (1). Subsequently, the response surface model using the MARS approximate function was constructed to emulate the daily DTVGM model applied to the Huaihe River Basin using the first 2000 samples and corresponding OBF values. In this study, the two parameters g_2 and *WM* were selected to generate the response surface model as shown in Figure 2. Figure 2(a) depicts the three-dimensional response surface from the output OBF and the input parameter g_2 and *WM*, while Figure 2(b) depicts the corresponding contour relationship. Figure 2 demonstrates that the ranges are limited to a relatively narrow interval when the OBF reaches the optimization value (i.e. the minimum value), and also demonstrates that the RSM can narrow the parameter ranges for the next optimization. In addition, the response surface model should be validated and verified before its application. The *k*-fold cross-validation and retest method provided by PSUADE was used in this study. Within the *k*-fold cross-validation, the original sample was randomly partitioned into *k* subsamples (k=500). Of the *k* subsamples, a single subsample was retained to be used as the validation data for testing the model, and the remaining *k*-1 subsamples were used as training data. The cross-validation process was then repeated *k* times. The statistical histogram of interpolation errors is shown in Figure 3(a). The scatter plot for simulation values from the response surface and calculated OBF values from the DTVGM model using the other 2000 samples is shown in Figure 3(b). It is evident that the relative error of less than 20% occurred for more than 90% of all samples, and the correlation coefficient was 0.96 between the simulation values from response surface and the output from DTVGM. The results demonstrate that the response surface model is acceptable and reasonable as a surrogate. The six parameters of the DTVGM model were optimized by searching the optimization values around the response surface. The minimum OBF was 0.073, and corresponding *WB* and *NS* were 1.002 and 0.856, respectively. The values of the six parameters were 0.387, 0.624, 0.039, 0.414, 0.092 and 0.460. The optimized parameters were applied within the DTVGM model using the data from 1991–2000, and the results showed that the RSM performed very well, and that it had a good accuracy, as shown in the Table 2. Therefore, the minimum data point on the calculated response surface was selected as the starting point for the next optimization. The parameter ranges were set as ±20% of the starting point values, and its maximum or minimum value was located within the original parameter ranges.

According to the work of Duan et al. [7], we can set the default values for the parameters within SCE-UA algorithm, i.e. m=2n+1, q=n+1, α=1, and β=2n+1, where n is the

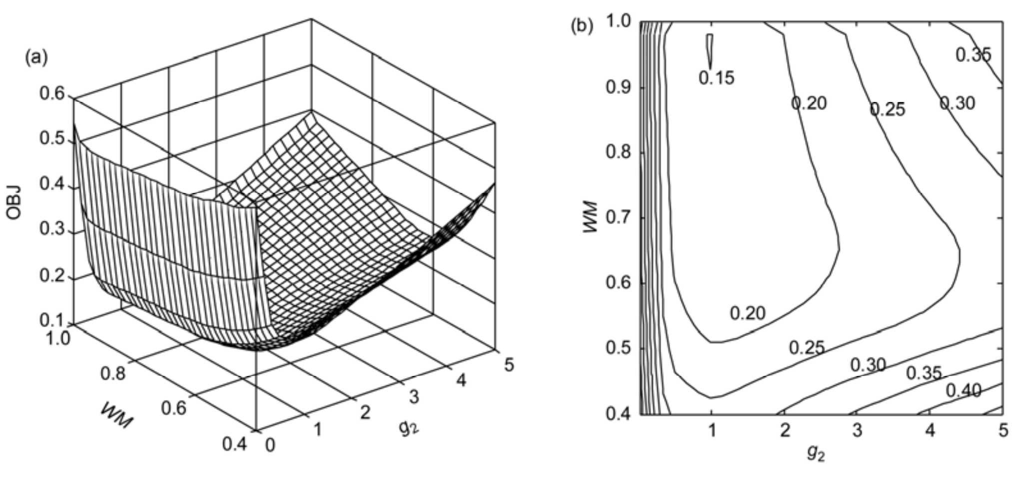

Figure 2 Response surface models (a) three-dimensional plot and (b) the corresponding contour plot.

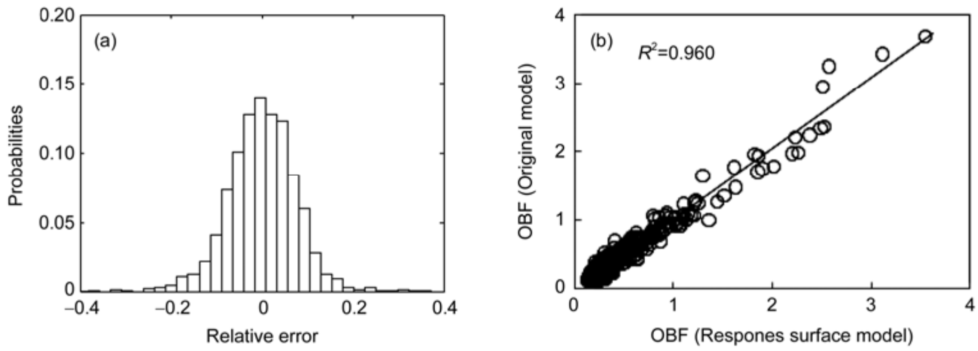

Figure 3 Results of the model validation and test (a) the statistical histogram of interpolation errors and (b) the scatter plot.

Table 2 Optimization results for the daily DTVGM model applied to the Huaihe River Basin

Method	g_1	g_2	*Kr*	Kaw	Wmi	*WM*	2003–2008			1991–2000		
							OBF	*WB*	*NS*	OBF	*WB*	*NS*
RSM	0.387	0.624	0.039	0.414	0.092	0.460	0.073	1.002	0.856	0.207	1.157	0.743
RSM-SCE-UA	0.405	0.502	0.031	0.406	0.106	0.547	0.062	0.999	0.877	0.125	1.009	0.759

number of parameters for optimization, m is the number of points within each complex, q is the number of points within each sub-complex, α is the number of consecutive offspring generated by each sub-complex, and β is the number of evolution steps taken by each complex. Hence, the only variable to be specified is the number of complexes p. A large complex is necessary for highly nonlinear problems. Therefore, in this study, the following parameters were kept constant: n=6, p=5, m=13, q=7, α=1, β=13. Two stopping criteria were used for the termination of the iterative process. The calibration process is terminated if one or more of the following criteria are satisfied: (1) the search stops when the algorithm cannot appreciably improve 0.01% of the output of the objective function over five iterations because of a very flat region of the response surface being reached; and (2) the search is terminated if the maximum number of iterations (10000) is exceeded.

In this study, the data from 2003–2008 was used to calibrate the model and to optimize parameters, while the data from 1991–2000 was used to validate the parameters. The optimization results are shown in the Table 2. Within the calibration, the OBF was 0.062, *WB* was 0.999 and *NS* was 0.877. Within the validation process, the OBF was 0.125, *WB* was 1.009 and *NS* was 0.759. Figures 4 and 5 show the fitting of the simulated and observed discharge for the cali-

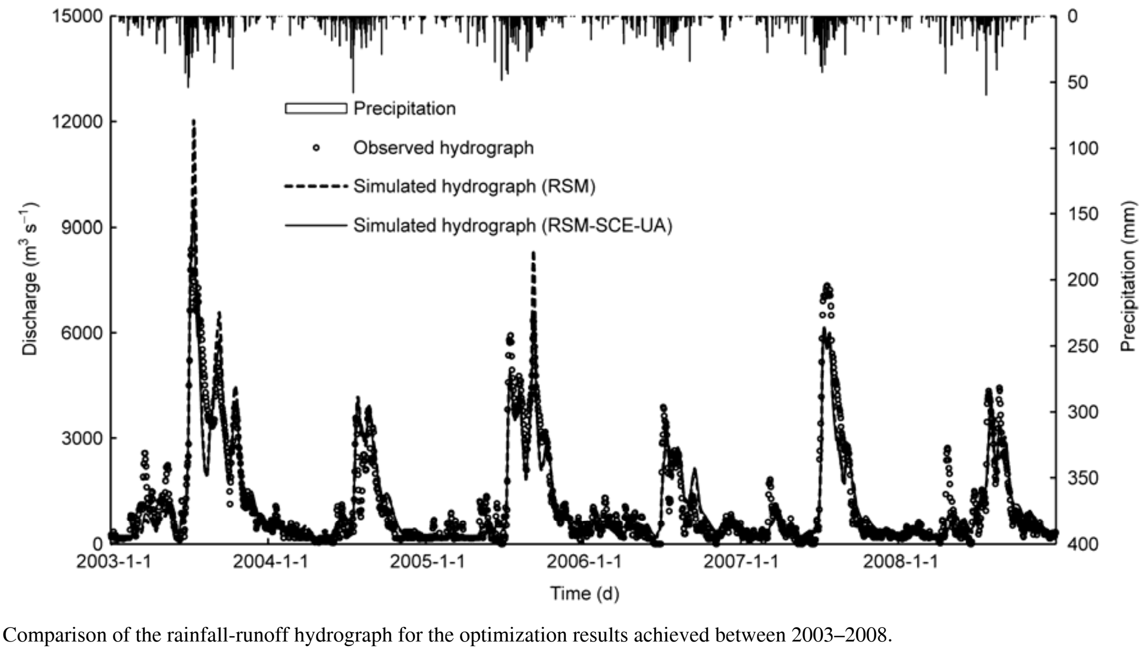

Figure 4 Comparison of the rainfall-runoff hydrograph for the optimization results achieved between 2003–2008.

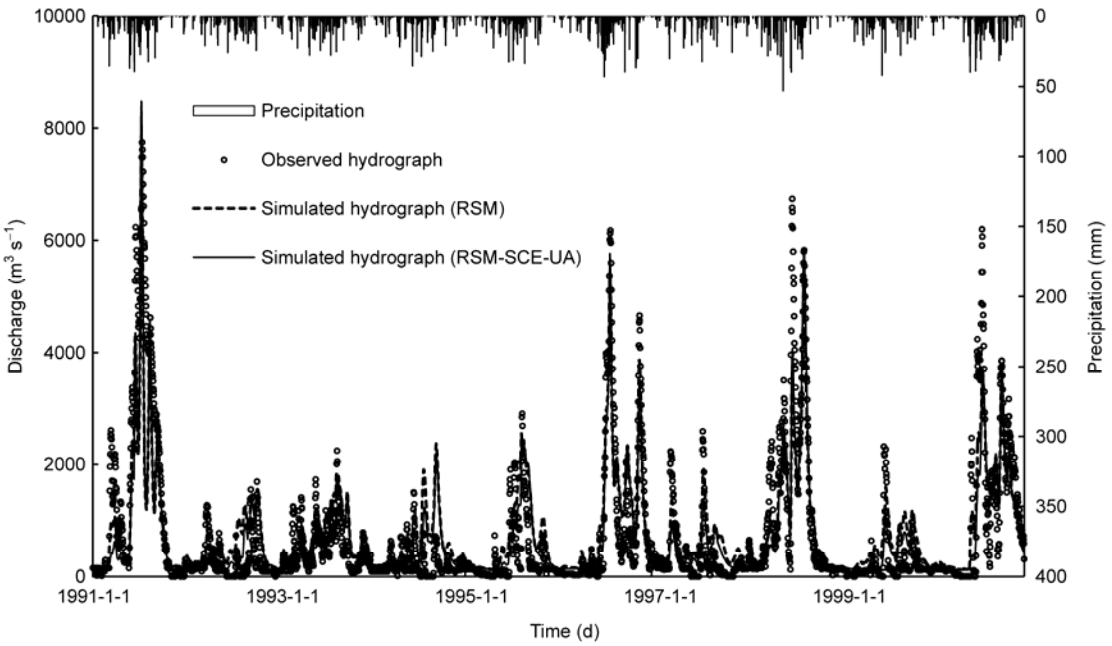

Figure 5 Comparison of the rainfall-runoff hydrograph for optimization results achieved between 1991 and 2000.

bration and validation processes. It is clear that the modeled discharge fits well with the observed discharge within the two study periods in the Huaihe River Basin. Table 2, Figures 4 and 5, demonstrate that the SCE-UA method (or RSM-SCE-UA) is capable of finding the global optimum parameter set for the DTVGM model. In addition, compared with the SCE-UA method, the RSM-SCE-UA method is an efficient global optimization method because it reduces computation costs and narrows the parameter ranges based on the statistical emulator approach. For a single SCE-UA method, the invalid parameter ranges may lead the search to be trapped in local optimums or cause the computation time to be much longer.

4 Conclusions

Parameter optimization is a necessary process within the application of distributed hydrological models. However, many automatic algorithms exhibit some disadvantages, such as being time-consuming, having an enormous computation cost, and being relatively difficult to implement in practice. In this study, the integrated method (RSM-SCE-UA) based on the statistical emulator approach and the SCE-UA method, is proposed for use within a hydrological model. Applying the DTVGM model on the Huaihe River Basin as a case study, the results show that the RSM-SCE-UA method can achieve efficient parameter optimization. The parameter optimization ranges can be determined rapidly by the RSM. In addition, the RSM method can output a narrow range if the method's effect on the output response is relatively large, or else it has a wide range. It is demonstrated that the narrowing range is crucial for determining the parameter ranges efficiently within the optimization processes. On the basis of the results of the response surface, the SCE-UA method is capable of finding a global optimum value for the DTVGM model without high computation costs. The integrated RSM-SCE-UA method would be capable of handling multi-parameter optimization for a complex distributed hydrological model with high parameter dimensionality.

The work was supported by the National Natural Science Foundation of China (*40901023*)*, the National Basic Research Program of China* (*2010CB428403*)*. We gratefully acknowledge the contribution of Dr. Ye Ai-Zhong* (*Beijing Normal University*) *and Dr. Zhang Yong-Yong* (*Chinese Academy of Sciences*) *for their help on the data processing and the DTVGM running. We also thank Prof. Charles Tong* (*Center for Applied Scientific Computing, Lawrence Livermore National Laboratory*) *for his help with the use of PSUADE, and Dr. Chen XiangDong* (*Chinese Academy of Sciences*) *for his help on the SCE-UA method.*

1 Song X M, Kong F Z. Application of Xinanjiang model coupling with artificial neural networks (in Chinese). Bull Soil Water Conserv, 2010, 30: 135–138, 144

2 Bahremand A, De Smedt F. Distributed hydrological modeling and sensitivity analysis in Torysa watershed, Slovakia. Water Resour Manage, 2008, 22: 393–408

3 Song X M, Zhan C S, Kong F Z, et al. Advances in the study of uncertainty quantification of large-scale hydrological modeling system. J Geogr Sci, 2011, 21: 801–819

4 Song X M, Zhan C S, Xia J, et al. An efficient global sensitivity analysis approach for distributed hydrological model. J Geogr Sci, 2012, 22: 209–222

5 Liang Z M, Li B Q, Yu Z B, et al. Application of Bayesian approach to hydrological frequency analysis. Sci China Tech Sci, 2011, 54: 1183–1192

6 Mariani V C, Coelho L D. A hybrid shuffled complex evolution ap-

proach with pattern search for unconstrained optimization. Math Comput Simulat, 2011, 81: 1901–1909
7 Duan Q Y, Sorooshian S, Gupta V K. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour Res, 1992, 28: 1015–1031
8 Chan K Y, Kwong C K, Luo X G. Improved orthogonal array based simulated annealing for design optimization. Expert Syst Appl, 2009, 36: 7379–7389
9 Lian K L, Zhang C Y, Shao X Y, et al. A multi-dimensional tabu search algorithm for the optimization of process planning. Sci China Tech Sci, 2011, 54: 3211–3219
10 Li D X, Liu W, Jiang J P, et al. Placement optimization of actuator and sensor and decentralized adaptive fuzzy vibration control for large space intelligent truss structure. Sci China Tech Sci, 2011, 54: 853–861
11 Omran M G H, Engelbrecht A P, Salman A. Bare bones differential evolution. Eur J Oper Res, 2009, 196: 128–139
12 Ge X S, Sun K. Optimal control of a spacecraft with deployable solar arrays using particle swarm optimization algorithm. Sci China Tech Sci, 2011, 54: 1107–1112
13 Afshar M H. A parameter free continuous ant colony optimization algorithm for the optimal design of storm sewer networks: Constrained and unconstrained approach. Adv Eng Softw, 2010, 41: 188–195
14 Herrera F, Lozano M, Molina D. Continuous scatter search: an analysis of the integration of some combination methods and improvement strategies. Eur J Oper Res, 2006, 169: 450–476
15 Reichert P, White G, Bayarri M J, et al. Mechanism-based emulation of dynamic simulation models: Concept and application in hydrology. Comput Stat Data An, 2011, 55: 1638–1655
16 Jin R, Du X, Chen W. The use of metamodeling techniques for optimization under uncertainty. Struct Multidiscip O, 2003, 25: 99–116
17 Lin Y S, Wang J, Wang X M, et al. Optimization of butanol production from corn straw hydrolysate by Clostridium acetobutylicum using response surface method. Chin Sci Bull, 2011, 56: 1422–1428
18 Ratto M, Pagano A, Young P. State dependent parameter metamodelling and sensitivity analysis. Comput Phys Comm, 2007, 177: 863–876
19 Sathyanarayanamurthy H, Chinnam R B. Metamodels for variable importance decomposition with applications to probabilistic engineering design. Comput Indust Eng, 2009, 57: 996–1007
20 Zhang H Z, Ming W W, An Q L, et al. Application of response surface methodology in surface roughness prediction model and parameter optimization (in Chinese). J Shanghai Jiaotong Univ, 2010, 44: 447–451
21 O'Hagan A. Bayesian analysis of computer code outputs: A tutorial. Reliab Eng Syst Safe, 2006, 91: 1290–1300
22 Duan Q Y. A global optimization strategy for efficient and effective calibration of hydrologic models. Dissertation for the Doctoral Degree. Arizona: University of Arizona, 1991
23 Box G E P, Wilson K B. On the experimental attainment of optimum conditions. J R Statist Soc B, 1951, 13: 1–45
24 Friedman J H. Multivariate adaptive regression splines. Ann Stat, 1991, 19: 1–67
25 Du Yuanyuan. The research of private enterprises credit scoring based on multivariate adaptive regression splines (in Chinese). Master Thesis. Changchun: Jilin University, 2007
26 Xia J, Wang G S, Lü A F, et al. A research on distributed time variant gain modeling (in Chinese). Acta Geograph Sin, 2003, 58: 789–796
27 Xia J, Wang G S, Tan G, et al. Development of distributed time-variant gain model for nonlinear hydrological systems. Sci China Ser D-Earth Sci, 2005, 48: 713–723
28 Wang G S, Xia J, Chen J F. A multi-parameter sensitivity and uncertainty analysis method to evaluate relative importance of parameters and model performance (in Chinese). Geogr Res, 2010, 29: 263–270
29 Sobol' I M. On the distribution of points in a cube and approximate evaluation of integrals. USSR Comput Maths Math Phys, 1967, 7: 86–112